INTRODUCTION TO MACHINE LEARNING
(ECEN 4122)

Time Allotted : 2½ hrs **Full Marks : 60**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve: **12 × 1 = 12**

Choose the correct alternative for the following

(i) Which of the following algorithms is best suited for a binary classification problem?
(a) K-means (b) Linear Regression
(c) Logistic Regression (d) PCA

(ii) Which of the following helps in preventing overfitting in a machine learning model?
(a) Adding more layers (b) Removing regularization
(c) Using a smaller dataset (d) Applying cross-validation

(iii) What is the purpose of hyperparameter tuning in supervised learning?
(a) To reduce the dimensionality of data
(b) To improve the model's performance
(c) To increase the training data size
(d) To reduce the number of classes in the dataset

(iv) Which of the following is an ensemble learning method?
(a) Decision Tree (b) Logistic Regression
(c) Bagging (d) K-means

(v) Which of the following is a boosting algorithm?
(a) Random Forest (b) AdaBoost
(c) K-means (d) Support Vector Machine

(vi) Which of the following is the key difference between bagging and boosting?
(a) Bagging reduces variance, boosting reduces bias
(b) Boosting reduces variance, bagging reduces bias
(c) Bagging increases the training dataset size
(d) Boosting uses parallel learning, bagging uses sequential learning.

(vii) In K-means clustering, what is the main objective of the algorithm?
(a) Minimize intra-cluster distance (b) Minimize inter-cluster distance
(c) Maximize accuracy (d) Maximize precision

(viii) What does the "K" in K-means represent?
(a) The number of iterations (b) The number of clusters
(c) The size of the dataset (d) The number of features

(ix) Which of the following activation functions is commonly used in deep learning models?
(a) Sigmoid (b) Mean Squared Error
(c) ReLU (d) Accuracy

(x) What is the main advantage of transfer learning?
(a) It requires large amounts of data
(b) It can quickly train models with limited data
(c) It is useful only for unsupervised learning tasks
(d) It is used to create completely new models.

Fill in the blanks with the correct word

(xi) The _____ algorithm is a popular classification algorithm that works by finding the hyperplane that best separates the data points into classes.

(xii) The _____ is an ensemble learning technique that reduces variance by training multiple models on different subsets of the data and averaging their predictions.

(xiii) In logistic regression, the _____ function maps the predicted values to probabilities.

(xiv) _____ is an unsupervised learning algorithm that identifies clusters of data points based on density.

(xv) The _____ layer in a Convolutional Neural Network (CNN) is responsible for reducing the spatial dimensions of the feature maps.

Group - B

2. (a) What is supervised learning? Discuss its key characteristics with examples. *[(CO1)(Understanding/LOCQ)]*

(b) Consider a data set, D={34, 36,42, 48}. Apply a suitable normalization technique to map the data to a new range 0-1. *[(CO1)(Apply/IOCQ)]*

(c) Explain over-fitting and under-fitting of a machine learning model. *[CO3][Understand/LOCQ]*

4 + 4 + 4 = 12

3. (a) Describe the concept of the bias-variance trade-off and its impact on model performance. *[CO1][Remember/LOCQ]*

(b) Construct the set of consistent hypotheses using Candidate Elimination algorithm for the data in the following table that describes the symptoms of the people and their Covid-19 test result. *[(CO1)(Create/HOCQ)]*

Inst.	Fever	Cough	Throat pan	Body pain	Covid 19
1	No	Yes	Yes	Yes	Poe
2	Yes	Yes	Yes	Yes	Positive
3	Yes	No	Yes	Yes	Positive

Inst.	Fever	Cough	Throat pan	Body pain	Covid 19
4	No	Yes	No	No	Negative
5	Yes	Yes	Yes	No	Positive
6	No	No	No	Yes	Negative
7	No	No	no	No	Negative

6 + 6 = 12

Group - C

4. (a) Explain linear regression and its assumptions. Compare and contrast Ridge Regression and Lasso Regression. *[CO2][Understand/LOCQ]*

   (b) Given the following data, build a linear regression model. Hence predict the score of a student who studies for 6 hours. *[(CO4)(Create/HOCQ)]*

Hours Studied	Exam Score
1	52
2	55
3	60
4	63
5	68

6 + 6 = 12

5. (a) Discuss the key advantages and limitations of using support vector machines (SVMs) for classification tasks. How do kernel functions enhance the performance of SVMs? *[CO4][Analyse/IOCQ]*

   (b) Using the provided house price dataset, implement a simple AutoML pipeline that selects the best model from a set of models (e.g., linear regression, decision tree, random forest) and tunes hyperparameters. Report the best model and its performance using MSE. *[CO6][Create/HOCQ]*

House_ID	Size (sqft)	Bedrooms	Age (years)	Price (lakhs)
1	1500	3	10	75
2	1200	2	15	60
3	1800	4	5	90
4	1100	2	20	55
5	1400	3	8	70

6 + 6 = 12

Group - D

6. (a) Define clustering and explain the differences between hierarchical and partitioning clustering methods. *[CO2][Understand/LOCQ]*

   (b) Using the given dataset, perform hierarchical clustering (Agglomerative). Create a dendrogram and use it to decide the optimal number of clusters. Report the clusters formed and evaluate their effectiveness. *[CO1][Analyse/IOCQ]*

Customer_ID	Income (k$)	Spending_Score (1-100)
1	40	60
2	80	90
3	20	30
4	50	70
5	60	50

6 + 6 = 12

7. (a) Compare and contrast various clustering algorithms (K-means, DBSCAN, Hierarchical, and GMM). Discuss the strengths and weaknesses of each algorithm in handling different types of data distributions. *[CO5][Evaluate/HOCQ]*

   (b) Using the provided dataset, create a clustering pipeline that integrates multiple clustering algorithms (K-means, DBSCAN, and GMM) to cluster patients based on their health data. Evaluate the results of each algorithm and justify which method is the most effective for this dataset. *[CO5][Create/HOCQ]*

6 + 6 = 12

Group - E

8. (a) What is Principal Component Analysis (PCA), and how does it reduce the dimensionality of a dataset? *[CO1][Understand/LOCQ]*

   (b) Explain the difference between feature selection and feature extraction in the context of dimensionality reduction. *[CO1][Analyse/IOCQ]*

   (c) Discuss the significance of feature scaling in dimensionality reduction. What are the common techniques used for scaling? *[CO1][Analyse/IOCQ]*

4 + 4 + 4 = 12

9. (a) Explain transfer learning and its importance in deep learning. What are the common approaches to applying transfer learning? *[CO3][Understand/LOCQ]*

   (b) What are recurrent neural networks (RNNs), and how are they used to handle sequential data? *[CO3][Understand/LOCQ]*

   (c) Discuss the advantages of using Long Short-Term Memory (LSTM) networks over traditional RNNs. *[CO3][Analyse/IOCQ]*

4 + 4 + 4 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	39.58	29.16	31.25